PERFORMANCE ASSESSMENT OF DIGITAL VIDEO TELECONFERENCING SYSTEMS

FINAL REPORT

CONTRACT DCA100-91-C-0031

Submitted to:
NATIONAL COMMUNICATIONS SYSTEM
ARLINGTON, VA

December 28, 1995

D E L T A I N F O R M A T I O N S Y S T E M S, I N C.
300 Welsh Road, Bldg. 3, Ste. 120
Horsham, PA 19044-2273
TEL: (215) 657-5270 FAX: (215) 657-5273

TABLE OF CONTENTS

1 INTRODUCTION

This document summarizes work performed by Delta Information Systems, Inc. (Delta) for the National Communications Systems (NCS), Office of Technology and Standards. The NCS is responsible for the management of the Federal Telecommunications Standards Program, which develops telecommunications standards, whose use is mandatory for all Federal departments and agencies.

This document is a final report for a Task Order on Contract DCA100-91-C-0031. The titles for the contract and Task Order are listed below.

- Contract DCA100-91-C-0031
  Development of Federal Telecommunication Standards Relating to Digital Facsimile and Video Teleconferencing

- Task No. 2
  Technical Work in the Area of Video Teleconferencing

- Subtask No. 3
  Video Teleconferencing Performance Measurement

Video TeleConferencing (VTC) is being widely deployed throughout the federal government, achieving great benefits in productivity and timeliness of decisions. The deployment has resulted from the maturity of video teleconferencing standards, the primary example being the H.320 series of ITU Recommendations. Video compression technology (Discrete Cosine Transform, Interframe Prediction, Motion Compensation) is the fundamental driver of the VTC revolution. The compression is based on digital technology, and the coded output picture is typically characterized by artifacts and distortions which are totally different from those in previous analog TV systems. New artifacts include tiling, jerkiness, mosquito noise, etc. As always, it is very important to measure and specify the end-to-end performance of any telecommunications system, and digital VTC is no exception. Since digital VTC is relatively new, the technology to assess system performance is very new, and the purpose of this document is to summarize the status of this work.

The report is divided into two main parts. Section 2.0 discusses the technical issues and background related to the assessment of VTC performance. Topics which are reviewed include digital vs. analog video, VTC applications, uses of assessment techniques, in-service vs. out-of-service testing, subjective vs. objective tests, analog vs. optical interfaces, combined video/audio assessment, single quality measure vs. multiple parameters, consistency, natural vs. artifical test patterns.

Section 3.0 of the report summarizes the present status of the work to develop procedures to assess the performance of VTC systems. The discussion is divided into four parts. Sections 3.1 describes a series of subjective tests for VTC which were recently performed by the ANSI/T1A1.5 standards organization. Sections 3.2 and 3.3

describe subjective and objective test standards respectively,which have been developed by domestic and international standards organizations. Addtional work to develop objective performance assessment measures is outlined in section 3.4.

Finally, conclusions and recommendations are included in section 4.0.

2 TECHNICAL ISSUES AND BACKGROUND

2.1 Digital vs. Analog TV

The advent of digital TV brings a number of problems in the measurement and specification of TV quality. For just the image portion of the TV system, there is a great deal of difference in the types of distortion that can be experienced.

For analog TV, there are only a few parameters that can be used to specify quality: frequency response, number of visible lines, noise and interference levels, multipath, frame rate, linearity, and various color distortions. It should be noted that for analog TV there was never developed a single quality measure that could be measured objectively. Rather, two other approaches were taken. One was to use subjective measure involving juries viewing TV pictures and rating them. The other was to make objective measures of various performance parameters, such as frequency response, and specifying the requirements for these parameters for each application. For example, the frequency response requirement would be higher for production quality TV than for the signal that is viewed in the home.

There are only a handful of analog TV standards in use throughout the world. Each standard specifies such quality measures as visible scan lines (which determines vertical resolution) and frame rate (which determines temporal resolution), so that these do not have to be measured if the standard being used is known.

This approach works well for analog TV because the displayed signal is basically a somewhat distorted version of the original. That is, the signal passes through amplifiers, modulators, receivers, etc. These devices filter, distort, add noise and other interference, but do not fundamentally change the nature of the signal.

Digital TV is fundamentally different from analog TV, in that the original signal is analyzed, and parameters that specify the signal are digitized and transmitted to the destination, where a replica of the original signal is reconstructed. Thus, the output TV signal could have a different frame rate, show a different part of the image, stretch or compress one of the dimensions, etc.

Furthermore, the video coder has a problem in the allocation of resources. The resource is the bitrate available to it, which is to be used to encode all aspects of the signal. If a large number of bits are needed to encode motion, this will leave fewer bits to code other aspects of the picture. Thus, motion in one area of the picture could affect the spatial resolution in another part of the picture.

Because modern video coders encode differences between adjacent frames, they try to ignore noise that may be present in the input signal, as coding this noise would waste the precious resource of channel capacity. One way to do this is to average several input frames of video, in order to minimize noise coming from the TV camera. If this is done, it may result in output video that is more pleasing to the eye

than the original video, but differs from it. Thus the traditional method of comparing the input and output video signals, and ascribing any differences to loss of quality, may give misleading results.

For analog TV, if the frequency response is measured using a test pattern, it can reasonably be assumed that the same frequency response will apply to any live video signal that happens to be applied to the system. For digital TV, this is no longer true. Because of the allocation of resources, the performance of the system depends to a great extent on the nature of the input video scene.

While there are a number of standards that have been developed for digital TV, including the codec functions, these standards do not limit the range of possible qualities that could be experienced. One reason is that the codec standards allow a very wide degree of freedom in the implementation of the standard, particularly at the encoder. Thus for Recommendation H.261, CIF or QCIF could be used, motion compensation is not required at the encoder, a minimum frame rate is not specified, etc. Furthermore, most VTC standards permit the escape to proprietary algorithms for the codec function. These proprietary coding algorithms could greatly change the range of qualities available, and could change the nature of the observed artifacts.

2.2 Applications of digital TV

2.2.1 General

Digital commercial television broadcasting, video teleconferencing, and similar technologies have become feasible as major applications of digital television as a result of the great improvements in digital video compression, digital video storage, and the improvements in digital communications. Digital television systems have become practical with the development of digital video standards. Standardization to assure compatibility is a necessity in order to reap the benefits of mass production without which the system is not practical economically.

Digital commercial television broadcasting is one application of digital television. A high definition digital television standard has been developed to satisfy the requirement to broadcast very high quality video (equivalent to 35 mm slides) and very high quality audio (CD quality) utilizing a communication channel. During the standardization process, it was found desirable for the broadcaster to be able to select various image formats and quality levels as well as the characteristics of the audio to be transmitted. The reasons were first, that not all of the material available warranted the very high quality capability and second, that as many as four video signals could be transmitted in a single communication channel if standard quality video were used. The latter is an important advantage. The new HDTV standard accommodates both of these, and other, features by defining performance levels. The standard specifies system parameters (video formats, audio service, etc.) to define a variety of performance levels. Since each performance level requires a specific data rate, a wide variety of services can be provided in the 19.2 Mbps (6 MHZ) broadcast channel.

Motion video teleconferencing is another form of digital television. It takes on many forms. The form which most commonly comes to mind is an exchange of video scenes of personnel, presentation material, writing on white boards, products, product applications, and displayed documents between two or more video teleconferencing rooms in each of which a group of conferees is participating. There are many related applications of digital video, other than conventional video teleconferencing, which serve probably an even much larger number of users. These applications include distance learning, training, engineering, data base retrieval, video phone, desk top video teleconferencing, conference broadcasting, and digital video disks.

The quality of the video communicated in a video teleconference is dependent on the purpose of the video teleconference. A high level of quality is a desired factor of course, because it must be adequate to permit the conferees to achieve the purpose of the video teleconference. But quality comes at a price: higher quality teleconferencing often requires more expensive terminal equipments and higher data rate communication circuits, also at higher cost. Image quality is also directly affected by channel data rate. Therefore it is prudent to match an acceptable quality of video with the purpose of the teleconference.

In both digital television broadcasting and in video teleconferencing, the quality of the transmitted and displayed pictures is determined by the level to which the following parameters have been implemented and faithfully reproduced:

Resolution
Gray Scale Rendition
Colorimetry
Motion
Frame Rate

The digital video communication system for broadcasting is almost always unidirectional. Video teleconferencing can also be characterized by the type of network configuration required for a specific application:

Point-to-point unidirectional
Point-to-point bidirectional
Multi-point
Broadcast

2.2.2 Commercial Broadcast / Entertainment Television

Several recent achievements have launched the era of commercial digital television.

- The completion of the United States Digital Television Standard for HDTV Transmission
- Successful completion of the Grand Alliance High Definition Digital

Television Prototype System

- Direct satellite television broadcast

2.2.2.1 U.S. HDTV Standard for Broadcast

The U.S. HDTV Standard for commercial broadcast television will provide outstanding features including noise free, high definition, wide screen, progressive scan video displays with low frequency enhanced surround sound Dolby audio.

The Advisory Committee on Advanced Television Services and the Advanced Television System Committee have been working for several years to develop a high resolution digital television standard for broadcast in the United States to replace the National Television System Committee (NTSC) system as the standard for the United States. In November of 1995 the standard was completed describing the Grand Alliance proposed digital high definition television system based on international ISO 13818 standard using the motion video MPEG-2 compression algorithm. The system was thoroughly tested in the laboratory and in the field. Only approval by the FCC is required to make it the U.S. broadcast television standard.

The U.S. HDTV standard defines a system which provides digital packetized transmission of audio, video, data, and a variety of other desirable signals all within a 6 MHZ channel (equivalent to 19.2 Mbps). Pictures, audio, and data can be transmitted in a variety of combinations. The packets contain the information necessary for the receiver to separate the packet types and to adapt to the various parameter levels used in the transmission of the pictures and sound. Stability is a key feature of digital systems and will probably most be appreciated in the resolution of the pictures, consistency of the colorimetry, and noise free display provided by the digital HDTV system.

The video signal can be transmitted in a variety of formats as shown below; one high definition signal or up to four standard definition signals at the broadcaster's discretion. Two aspect ratios are provided; a new movie like 16:9 and the present 4:3. Picture rate is also a variable depending on the motion content and source of the pictures.

FORMAT (Pixels)	ASPECT RATIO	VERTICAL RATES (Pictures/second)
1920 x 1080	16:9	60 interlaced and 30, 24 progressive

1440 x 1080	16:9	60 interlaced and 30, 24 progressive
1280 x 720	16:9	60, 30, 24 progressive
960 x 576	16:9	
720 x 576	16:9	
704 x 480	16:9, 4:3	60 interlaced and 60, 30, 24 progressive
640 x 480	4:3	60 interlaced and
352 x 288		60, 30, 24 progressive

The audio system does not follow the international standard (ISO 13818) but is based on the Dolby AC-3 system. It can provide combinations of the following features.

Two channel stereo
Five channel surround sound
Low frequency enhancement channel
Program voice transmitted separately from effects and music
Second language program channel
Dolby
Commentary channel
Hearing impaired
Voice overs
Narrative for the visually impaired
Emergency messages

Other features provided by the U.S. standard include:

Two data channels
Program subtitles
Limited (controlled) access
Program guide
Virtual channels
Network data
Error correction

Initial digital HDTV broadcast transmission will probably occur in 1996 with commercial broadcasting beginning in 1997 on a simulcast basis.

2.2.2.2 Direct Satellite Broadcast

Several of the features described above to be provided by the U.S. HDTV broadcast standard have already been implemented in the several existing direct satellite to the home television systems. Principal among these is the fact that digital,

packetized, transmission is used. The use of digitized video to capitalize on the powerful compression algorithms permits a large number of channels to be accommodated by the digital transmission system. Color consistency and noise free reception are noticeable features.

2.2.3 Video Teleconferencing

Video teleconferencing generically satisfies the need for point-to-point visual communication. The function to be performed and material to be transmitted and displayed varies widely with the specific application. Each specific application requires the system parameters to be implemented to its own acceptable performance level. As a result systems exist which are very similar in architecture but each is implemented with uniquely specified levels of parameter performance. For example, a video teleconferencing system for executive use differs from one for engineering use in the level of the system parameters. In fact, a range of systems applications exists in addition to these two. The distinction is made not to order the importance of the application but rather the difference in performance of the various functions and the parameter levels required as described below. Standards have been prepared which define the ranges for each parameter for given system types to assure compatibility.

In video teleconferencing it is necessary to synchronize and minimize video and voice delay, such as transmission and switching delay among conferees.

2.2.3.1 Executive Level Video Teleconferencing

The term executive level video conferencing is used here to denote the video teleconferencing systems with the highest level of features. The quality of this video system is more demanding than other forms of video teleconferencing discussed later, because of its functions, high levels of resolution, gray scale, colorimetry, motion capability, and accommodating multi-users.

Executive video teleconferencing systems are most commonly implemented in a bidirectional point-to-point configuration. A system involves at least two user terminals connected through a bidirectional communication circuit. Each terminal displays, to the local users, the scene transmitted by the distant terminal. Often the terminal is embellished to also provide a preview display of the local scene prior to transmission. The video consists of images of the users, text, graphics, scenes, and pictures related to the topic under discussion. Audio may be simultaneous from both locations but more often is controlled to permit only one user to speak at a time. Unidirectional point-to-point video teleconference is comparatively unusual. Applications are described later.

Executive level video conferences are usually satisfied by "NTSC level" resolution or, in digital terms, 640 x 480 pixels. Displays of the conferees' faces, the conference table, and scenes, products, etc., transmitted at this level of resolution are acceptable. Resolution and color are prime system parameters. Equally important is motion portrayal because of its effect on channel data rates. The motion required by

typical executive level video teleconferencing scenes can be adequately supported by motion frame (information) rates achievable using channel transmission rates between 384 Kbps and 1544 Kbps over readily available fractional or full T-1 transmission channels. The digital video compression implemented in terminals of this type allows the 384 to 1544 Kbps transmission of 640 x 480 pixel frames in color or in gray scale at a motion frame rate with adequate capability to satisfy "executive" requirements.

At the executive level, the textual and graphical material for use in the video teleconference are often prepared specifically so that the resolution of conventional systems provides an excellent display. However, full page textual and graphical material requires higher resolution; for example, facsimile level resolution of 2200 x 1280 pixels. The video teleconferencing standards protocols as implemented in video teleconferencing equipment provide the capability to transmit random full page text and graphics at a higher resolution by transmitting it as data, sharing the transmission stream with the video signal when required. The receive terminal stores the signal and refreshes the text or graphics display from the digital memory thereby minimizing the load on the communication channel. Computer data files such as spread sheets are accommodated in a similar manner directly from the computer communication port via the video teleconferencing terminal.

Multi-point video teleconferencing has become very popular as a result of recent multi-point control equipments made possible by the development of multipoint control standards. This technique permits many users to participate in a single conference. Techniques for switching among users have also been improved. Selecting the user whose video will be transmitted to all locations is at the heart of multi-point video teleconferencing. In multipoint systems, all users are connected via a control function which selects the video signal to be displayed to all users. Switching can be manually controlled by the conference manager or can be on a voice controlled basis, user request basis, or other basis. The user terminal may also provide a preview display of the local picture to be transmitted. The network must consist of bidirectional or reversible circuits since each user must receive and, at some time, transmit video signals. The time delay between the conclusion of one transmission and the beginning of the next transmission must be considered to prevent two users from attempting to transmit simultaneously. For example, in networks configured using satellite channels, the total delay could be up to a substantial fraction of a second. Care must also be taken so that the voice and video paths among all users do not introduce unequal delays causing an annoying lack of lip sync. Audio signals may be connected through a bridge so that any participant may speak at will, but are more commonly selected as a part of the video signal selection.

2.2.3.2 Engineering Level Video Teleconferencing

Engineering video teleconferences may be identical to executive video teleconferences when the purpose of the conference is to discuss technical issues among a group of engineers. Resolution, gray scale, color, and motion requirements (and therefore also data rate) would be the same in this application.

A video teleconference for engineering purposes generally does not require "people pictures" as the principal information conveyed because attention is focused on the subject of the teleconference which may be engineering drawings, sketches, and comparatively low motion scenes. The video is often transmitted in one direction via a unidirectional transmission system. It is possible that, for a specific application, a return video channel is necessary requiring a bidirectional circuit. Audio requirements are generally bidirectional.

Drawings require only bi-level video, but if scenes or photographs are part of the teleconference material, gray scale or color images are required. Seldom is "good" motion portrayal required (generically speaking). This combination of features may require a high resolution display such as facsimile level, or higher, so that the details of the displayed material are clear. Even though resolution requirements may be high, the fact that motion is minimal, permits the video data to be transmitted at a comparatively low data rate as long as the time required for the development of a new data display is not annoying. Practical data rates are 128 Kbps such as can be provided by basic rate ISDN circuits. Digital image storage is required at the receive end to permit display refresh from the receive terminal rather than from the transmit terminal via the communication channel. Bidirectional voice is generally required.

2.2.3.3 Broadcast Video Teleconferencing

Broadcast video teleconferencing is a special form of multi-point video teleconferencing in which switching is not required. A specific terminal is selected to be the source of the video and audio which are transmitted to all other terminals. The secondary terminals generally cannot respond with either video or audio. A classic application is the corporate chairman addressing all employees in diverse locations. The technological key is to provide a single video signal to many points. Often a circuit to each point is provided or satellite transmission may be employed. In other cases a single circuit routed through all points with drop and forward equipment is utilized; e.g., on fiber optic SONET ring systems.

The video quality required for this application also depends on its specific purpose. For the example cited, it is the same as the executive teleconference.

2.2.3.4 Room/Desk Top Video Teleconferencing

Video teleconferencing can also be categorized as to the location of the video teleconferencing equipment and the conferees. The parameters of the video teleconferencing system and the communications channel used will also vary.

Room video teleconferencing is akin to the executive level of video teleconferencing. A fairly large size room is dedicated to the teleconferencing function; lighting, acoustic treatment, sound enhancement, large screen displays, remote control for camera pan, zoom, tilt, remote control for the teleconference and transmission

functions, convenient location of supplementary equipment, conference table, and comfortable seating arrangements for a number of conferees and observers are all carefully planned and incorporated. Document scanners and PC's may be included for document and data file transmission via an auxiliary data channel within the video data stream.

The material for the conferences, the video and audio parameters, and the transmission channel are essentially the same as those for the executive level video teleconference;

- scene resolution - 176 x 144 pixels
- document resolution - 2200 x 1280 pixels
- gray scale - full for scenes, generally bi-level for documents (although color and gray scale may be required)
- color - full for scenes
- data rate - 384 to 1544 Kbps, bidirectional, broadcast, or multipoint
- data files - transmitted as data

Desk top video teleconferencing is different in purpose and in implementation from that conducted in a video teleconference room. The conference terminal is generally located in an office or a small conference room and is intended to be used by one, or at most a small number of conferees. Discussions between single conferees at each end of the system is a common application. It is akin to the engineering type of video teleconferencing. Most important, excellent motion portrayal is not a requirement so that high data rate transmission is not required. The desk top terminal equipment is comparatively small and, in fact, can be a software and hardware implementation within a personal computer. Large screen displays, remote pan, tilt, zoom control, and remote teleconference control are not required since all controls are at the conferee's finger tips on the desk top. Although small physically, it incorporates the more important features of the higher level systems. The display is presented on a video monitor, often the computer monitor, a single camera televises the conferee within a fixed viewing area. The main transmitted material are documents, for which a scanner may be provided, and data files, which are accessed directly from the computer. The system is often enhanced with a keystroke program by means of which the conferees at either end of the system can manipulate data from a file which is displayed on their screen. The material for the teleconference can be generated in the PC and stored on the hard drive. The files can be randomly accessed for use during the video teleconference.

- scene resolution - 640 x 480 pixels
- document resolution - 2200 x 1280 pixels
- gray scale - full for scenes, generally bi-level for documents (although color and gray scale may be required)
- color - optional
- data rate - 128 Kbps, bidirectional
- data files - transmitted as data

2.2.4 Distance Learning and Training

Distance learning and training can be considered specialized forms of video teleconferencing. They are generally characterized by the fact that the video transmission is unidirectional. Video information flows in one direction; from the instructor/trainer to the student(s). Audio is more often bidirectional. Provision for students to question the instructor is highly desirable as everyone who has engaged in teaching activities is aware.

The video signal must provide a very clear presentation of the material under discussion. Every picture presented must be clear, unambiguous, and easy to assimilate. Other forms of video teleconferencing may utilize the accompanying voice narrative to aid in clarifying shortcomings in the images presented, but for distance learning and training the images must stand on their own while the narrative explains the principles involved. This places exacting resolution, gray scale, and colorimetric requirements on the performance of the video communication system. Motion requirements depend on the specific application of the distance learning system. If the system is multi-purpose, good motion portrayal will be required at some time. The resolution requirements are eased by careful zooms or close-ups rather than providing a more complex system. All distracting artifacts such as motion flicker must be carefully avoided. Great effort is generally devoted to preparing a quality presentation (and is an economically sound trade-off).

As an example, a major automobile manufacturer utilizes video teleconferencing via satellite simultaneously to a nationwide audience of actually hundreds of automobile sales and repair personnel to describe new models or specific problem items on a regularly scheduled basis. The video is clear and properly scaled to show the points of interest in real time or in enhanced motion for clarity. High resolution or unusual video is not required as a result of careful planning and production of the conference material and is satisfied by executive video teleconference levels. The voice accompaniment is synchronized with the video. Ample time is provide for audience questions. The return audio path is via dial-up telephone circuits into an audio bridge co-located with the instructor.

2.2.5 Video Phone

Perhaps the most discussed form of digital video is the video phone. The video phone provides personal image and voice transmission between terminals (video phones) located in the home utilizing the public switched telephone network. Video phones have been forecast by various manufacturers for over 30 years. For several reasons it has not become popular: lack of broad-based interest, use of analog technology, the cost of implementation, unavailability of practical/low cost video memory, and inadequate video compression technology to utilize existing communications circuits. But the world has turned since then and low cost memory is now available, compression technology has improved by orders of magnitude, component cost has dropped due to mass production, a major advance in modems utilizes existing PSTN communication circuits, and, most important, the public mind set

has been modified to expect implementation of the promises of the past.

Not to be forgotten, a platform for video phone now exists in most homes and offices which can provide this capability; namely, the home computer. While the home computer is not the production terminal implementation of the future video phone, it may greatly foster public interest (e.g. similar to booming Internet activity). Another major change, perhaps more important than the technological changes, is that international standards are being developed to provide a unifying, compatible, protocol for video phone. (The risk to manufacturer's investments has been greatly decreased.)

The quality of imagery required for the video phone application is quite nominal. The main requirement is that the viewer be able to identify the transmitting subject and determine what he is doing as a motion requirement. Resolution, gray scale, colorimetric, and motion specifications for these systems are modest but must be met if video phone is to be successful. Standards under development define a protocol for transmission of picture phone images at rates up to 28.8 Kbps. Motion information may be well below the 30 frame per second display refresh rate and still provide satisfactory performance for video phone applications. QCIF (176 x 144) and CIF (352 x 288) can provide extended quality picture phone displays, if required, using basic rate ISDN circuits at 128 Kbps or a single B circuit at 64 Kbps.

The video phone era is probably about to enter the scene in a grand manner.

2.2.6 Data Base Retrieval

Data base retrieval requirements are similar to engineering teleconferencing requirements except that video transmission is generally unidirectional. A reverse data channel is also required to access the data system. The resolution level, gray scale, and color capability must be adequate to support the material being accessed. Data base applications are extremely varied. For example, a data base may consist of specially formatted textual and graphical material such as preformatted stock market reports, weather forecast, etc. These are low resolution, bi-level, static images and can be conveyed as data via low data rate circuits. Other data bases may contain letter size documents (e.g., medical and legal records) and therefore require considerable resolution for unambiguous display. If adequate time is permitted for transmission, low data rate channels may be used. Data bases with photographic material are common in, for example, real estate sales. The addition of color and/or gray scale increases the amount of data to be transmitted but, if there is no motion, can be conveyed over low data rate circuits. If the transmission delay is annoying, higher data rate circuits can be used. X-ray data bases require excellent video with high resolution and extended gray scale. Although the resolution for these requirements may be high and the gray scale extended, no motion is required so that comparatively low data rate transmission circuits can be used depending on the receiving personnel tolerance to transmission delay. The higher resolution images require higher resolution terminals with display refresh. In the news media, data bases containing motion clips are widely used. For these applications, conventional resolution, gray scale, and color will suffice but the

motion requirements dictate considerably higher data rate circuits.

2.2.7 Digital Video Disks

Digital video disks using the MPEG compression algorithms have become very popular. Their content ranges from feature motion pictures to the upper quality video games. The video quality is dependent on the material recorded which to date is the digitized NTSC version of the motion pictures and VGA video games. However the capability should become equal to, and be compatible with, that of the HDTV standard. The disks are inexpensive and very tolerant of handling.

The sudden increase in popularity of the video disk is due to the CD ROM implementation and the fact that many homes already have a disk player; namely, the CD ROM drive in their home computer.

2.3 Uses of Assessments

As with analog measurements, there are a number of potential uses for the quality assessments of digital TV systems. Some of these uses are described in the following sections.

2.3.1 In-place Systems

One use for quality assessments is to determine the possible degradation over time of a digital TV system that has already been installed. By making measurements at installation and periodically over a long period of time, subtle degradations in performance can be detected. In addition, the nature of the degradations can lead to a diagnosis of the location and nature of the problem. For example, it can point to the coder, decoder, or network, and responsibility can be assigned. Of course in a digital system most failures are catastrophic, not gradual. In order to make this type of measurement, the two ends will likely be at different locations.

It should be noted that the only application of the draft ANSI standard that is approved is this type.

2.3.2 Procurement Requirements

The existence of objective picture quality measures makes it possible for procurement agencies to specify minimum performance levels. They can be sure that they will acquire systems that will provide acceptable video quality, and that manufacturers do not take short cuts to reduce cost, such as fail to use motion compensation. Thus units that only meet the letter of the mandatory standards, but do not fully utilize them, can be eliminated from consideration. This is particularly important for government procurements, where the fairness of a procurement must be apparent.

The requirements can be expressed as minimum values on individual performance values, or as an overall quality measure.

2.3.3 Acceptance Tests

Once the VTC has been procured to a quality standard, it must be tested to insure that it meets the requirements.

For this type of measurement test sequences can be used, and the test can be performed in the laboratory, with both terminals at the same location, with a communications loop if required. It is also important that consistent results be obtained, so as to protect the interests of both the vender and user of the equipment.

2.3.4 System Comparison

In some cases a user is considering the acquisition of a video communications system, and wishes to purchase the system that provides the best performance for the price. He may be faced with several possible systems, and their performance and price may be quite similar. This means that measurements must be quite sensitive to small differences between the performance of various systems. In some cases the user will be interested in only an overall quality measure, while in other cases he will be interested in individual parameters.

This is a politically difficult application, since each vendor will claim that the measurement procedure is biased against his equipment.

2.3.5 Determining Bandwidth Requirements

A user may wish to determine what type of network he needs to support his teleconferencing application. Specifically, he may wish to know what transmitted bit rate is needed to provide acceptable performance. If there are published minimum performance levels for different applications, such as room teleconferencing, he can try his proposed video equipment at different bit rates to determine the least expensive network that will meet his needs.

2.3.6 Determining Error Sensitivity

Different networks may offer different levels of error performance. In order to evaluate the sensitivity of codecs to transmission errors (or lost cells in an ATM network), bit errors can be simulated. Then the objective performance measures can be used to determine the quality of the output image.

2.3.7 Evaluating Interoperability

It is desired to evaluate the degree of interoperability of video codecs supplied by different venders over different communications networks. In many cases, a particular combination will either work (provide an output picture) or it won't. However, it may be that in some cases, while a picture may be provided, the quality of the picture is not as good as that when the same vendor's equipment is on both ends (perhaps because of the use of proprietary algorithms). A method to quantify the loss of quality when different equipments interoperate is needed. Note that the quality loss may be different in each direction in this case.

2.4 In-service vs. Out-of-service tests

In some applications, it may be desirable to perform objective tests during the actual transmission of live video signals. This could be the case for the application of measuring the degradation of in-place systems. However, there are a number of problems with this approach.

One problem is that there is no control over the content of the video scene that is being transmitted during the test. This means that the image could be a blank screen, which would provide almost no information about system performance, or a rapid sequence of scene cuts, which would over-stress the codecs, and give misleading results.

Another problem is that the scene content will vary from one test to another, making it very difficult to obtain consistent results from one test to another, and from one test implementation to another, no matter how sophisticated the test algorithm is. Inconsistent results will give rise to doubts about the accuracy of the measurements.

In addition, with live signals there is no way to tag input frames so that they can be identified at the output. Signals added to the blanking intervals will not be transmitted to the destination, and signals added to the visible portion of the signal will interfere with the normal operational use of the system. As will be seen, tagging frames is useful for defining the start and end of the test, and providing a simple and reliable method of measuring transmitted frame rate and jitter in frame rate, as well as time delay.

On the other hand, with an out-of-service test, the input signal can be precisely controlled, input frames can be tagged so as to be identifiable at the output, an tests

can be run over exactly the same frames on each test. All of the applications described in Section 2.3 can be performed out-of-service, with the possible exception of in-place systems, but there will usually be periods when test signals can be inserted, such as at the beginning or end of each conference.

2.5 Subjective vs. Objective tests

Subjective measures of video quality can be made using juries and accepted testing procedures. In fact, the T1A1.5 used these procedures to establish a baseline against which to compare objective measures. Reference A provides the guidelines for such measurements. See Section 3.1 for details.

Subjective tests have a number of problems. One is that they are very costly. They require a carefully constructed facility, a large number of personnel, and a long time. It took the extensive resources of the T1A1.5 committee many months to perform one series of tests that compared 25 different "systems". While this cost might not be excessive for some applications, for most it could not be borne.

Another problem is the time delay before the results of the measurement become known. This could be days or even weeks, rather than the minutes that are required for some applications.

Subjective measurements require extensive juries of observers to judge the quality of the video. This is not exciting work and cannot pay very well. There may be great difficulty in recruiting people for this work.

Because of the human element, there will always be differences of opinion about the quality of the video image. This leads to inconsistent results from measurement to measurement, which can be reduced only by using very large juries.

In general, juries can only give overall impressions of video quality. If measurements of the various parameters that contribute to video quality are desired, then experts who are familiar with digital video artifacts would be needed.

For the above reasons, objective measures must be developed for measuring video quality, with the possible use of subjective tests to provide a baseline.

2.6 Analog vs. Optical Interfaces

A serious problem exists if the system to be tested has an integrated camera, and does not possess a standard analog (or digital) video electrical input interface, such as NTSC or PAL. In this case there is no point at which a test signal can be introduced, or at which the characteristics of a live input signal can be measured. This type of system could be tested only by establishing a standard scene or set (such as the "mobile with calendar" scene that has been used to test codecs).

In addition, if the system to be tested has an integrated display, and does not possess an electrical interface, it would not be possible to directly measure the parameters of the output signal. An unattractive approach would be to have a camera viewing the output display to produce an electrical signal that could be analyzed.

2.7 Video, Audio, and Combined Assessments

The primary focus of this document is the performance assessment of digital video teleconferencing systems. However, it should be noted that audio plays a major role in the overall performance of teleconferencing systems. In fact, it is generally recognized that audio is the fundamental cornerstone to the overall performance in an audiovisual system. In general, without audio, it is impossible to have any teleconference. In addition, tests have shown that the audio quality greatly influences the perceived overall audiovisual quality of the teleconference.

Digital audio coding has been practical and commercially available for a long time - much longer than digital video coding. Consequently, the procedures for testing digital audio coding systems are much more mature than are for testing digital video systems. Techniques for audio subjective testing are extremely well defined. It should be noted that there is a major distinction made for the performance of audio systems which must handle a wide range of signals such as music as opposed to those which code only speech signals. For example, the G.726 (16 Kbps) and G.723.1 (5.3/6.3 Kbps) speech coders are primarily designed to handle speech as opposed to a wide range of audio signals.

The procedures for subjective testing of speech coders are extremely well defined. The techniques include tests for different languages, genders, background noise, and various error rates for differing numbers of tandemed circuits.

Objective testing of digital speech/audio coding is much more mature than objective testing of digital video. Nevertheless, it is far from being considered to be equivalent to subjective testing. Consequently, subjective testing of digital audio/speech systems is most common, and work continues on the development of techniques to objectively assess the quality of digital speech coders.

Since teleconference systems are gaining widespread acceptance, there is interest in assessing the performance of the integrated multimedia terminal. Work in testing integrated audiovisual terminals is still in its infancy. Issues to be addressed include audio/video synchronization and relative audio/video bit rates.

The ITU is developing two standards which assess the overall performance of an audio visual terminal. See Section 3.2.2 for a more detailed discussion. Recommendation P.920, entitled AInteractive Test Methods for Audiovisual Communications@, defines subjective/objective test methods for measuring the ability to perform interactive conversational tasks by means of multimedia terminals. Recommendation P.AVQ, entitled ASubjective Assessment Methods for Global

Audiovisual Quality Evaluation in Multimedia Applications@, defines non-interactive subjective assessment methods for evaluating the global quality of digital video with accompanying audio for audiovisual applications such as videotelephony, videoconferencing and storage/retrieval. The Recommendation describes the methods which are suitable for evaluating the following quality aspects:

1) the global effect of coding impairments, taking into account the interactions between audio and video degradation,
2) problems related to lack of synchronization between the two signals,
3) the impact of transmission errors on such video systems.

2.8 Single Quality Measure vs. Multiple Parameters

For some applications, there is some desire to obtain a single quality measure for digital video systems. One such application might be a system comparison, where it is desired that systems be ranked. This would be difficult to do if one system had better spatial resolution while another had better motion response.

For analog TV systems, there are no objective measures of overall video performance. This is because users are comfortable in specifying the individual parameters that contribute to overall performance.

For most applications, having individual performance parameters is superior to a single overall measure of performance. This is because for each user there are different aspects of the video signal that are important to him. For example, for surveillance applications the detail of the image is more important than very fast motion response. Conversely, for videophone applications, good motion rendition might be more important than a large amount of detail in the image. When individual quality parameters are presented, each user can give them a weighting that is appropriate for his application.

For each application, there could be several levels of service defined. Each level of service would have a list of requirements in terms of the individual performance parameters. This would be an aid to writing specifications for procurements, and testing for acceptance.

While there have been several attempts at calculating an overall quality measure from individual measured parameters, these have not been entirely successful. The correlation between the overall measure and the subjective measure obtained from extensive tests has been poorer than expected. (See Section 3.1)

While the possibility of a single objective quality measure should not be dismissed completely, this should be the final stage of the development, after individual quality parameters have been carefully defined and tested.

2.9 Consistency

Even with well-controlled test patterns, and frame tagging that allows running each test over the exact same frames, there is still a serious problem in obtaining consistent results from test to test. This is because the encoder may have its own irregularity or periodicity. For example, the encoder may provide Intra mode refreshing at some rate. If Intra refreshing is done a frame at a time (allowed by all standard algorithms, and required by some, such as MPEG), then the test results may depend upon how many Intra frames are transmitted during the test sequence. For the H.261 and related standards, it would only be necessary to transmit one Intra coded frame every 132 coded frames. If frames are coded at 10 frames per second, an Intra frame would occur only once every 13.2 seconds. Therefore a test integration period less than 13.2 seconds may or may not contain an Intra frame. Thus some tests would give different results from others. In order to obtain consistent results, the integration period would have to be much larger than the periodicity of the Intra refresh rate.

2.10 Real Scenes vs. Test Patterns

There is a question as to the type of input material that should be used to perform various tests.

The use of real (natural) scenes has the benefit that they stress the codecs in realistic ways, that it is almost impossible to do with artificial test patterns. This is not to say that uncontrolled live signals should be used. As described in Section 2.4, in almost all applications it is possible to used canned sequences, in order to obtain the best possible consistency.

On the other hand, artificial test sequences are useful in measuring a specific aspect of performance, such as resolution or motion rendition.

A possible compromise is to use test scenes that are made up largely of canned natural scenes that are typical for the intended application, and to insert into these scenes small windows of artificial test patterns, in order to measure specific performance parameters. In addition, fiducial marks can be inserted to simplify spatial registration. Time codes can be used to identify each frame. These permit a clear definition of the start and finish of the test, and simplifies the measurement of the average transmitted frame rate and frame rate jitter. The key is to make the windows small enough so that they do not themselves provide a significant load to the codec, but large enough so that they are well-defined even at the lowest bit rate.

3 WORK STATUS

There is a great deal of activity, on a worldwide basis, to develop technology and standards to assess the performance of digital video communication systems. The performance of any TV system can be measured using two fundamentally different approaches and technologies--subjective and objective. Subjective tests employ the human eye as the ultimate measuring device (therefore very costly and time consuming), while objective tests rely solely on instruments. Since the characteristics and impairments of digital video are fundamentally different when compared to analog TV, the technology for measuring digital picture quality (both subjective and objective) must be updated. As would be expected, the work to subjectively assess performance of digital video communication systems is far more mature than the technology to objectively measure picture quality. One basic reason for this situation is that all candidate objective measures must be validated by correlating them with ground truth subjective test data.

As usual it is very costly to measure performance by subjective means, so there is a strong motive to develop new objective measures. Although considerable effort has been focused on this topic, much work remains to be done.

The purpose of this section is to summarize the status of the work to develop performance assessment techniques for digital video teleconferencing systems. The work is divided into four parts; (1) subjective tests performed by ANSI, (2) subjective test standards, (3) objective test standards, (4) additional objective performance assessment studies.

3.1 ANSI Subjective Tests

During 1994, the ANSI sub-committee responsible for video quality measurement (T1A1.5) conducted a series of subjective tests on a variety of teleconferencing codecs. The tests were performed in accordance with ANSI contribution T1A1.5/94-118R1, "Video Performance Standard Subjective Test Plan". Tests were performed at three different and widely separated laboratories, one of which was Delta. The results of processing 25 different test scenes through 25 different codecs (Hypothetical Reference Circuits - HRCs) were recorded on D2 digital tape. Twelve analog test tapes were then produced with a random ordering of 64 or 65 of the original scenes followed immediately by the same scene processed through one of the HRCs.

At each lab, at least 30 observers were recruited who were not video experts, some of which had some experience with video conferencing and some of which did not. These observers were divided into three teams (red, green, and orange). Members of the same team at the different labs saw the same four tapes with the same HRCs on them. A few of the HRCs were assigned to two or three different teams, and so had more observers.

At each lab a room was made available for conducting tests in accordance with

CCIR 500-5. Up to three observers sat for each test session, which would last one or two hours. Each observer saw all four tapes assigned to his team. For each HRC/scene combination, the observer was asked to indicate whether his opinion of the difference in quality between the reference and processed scene was: Imperceptible, Perceptible but not Annoying, Slightly Annoying, Annoying, or Very Annoying. These opinions were later converted to the numbers 5 through 1 respectively.

The result of the above was a set of at least 30 subjective opinion scores for each of the 625 (25 x 25) HRC-scene combinations. Contribution T1A1.5/94-148 "T1A1.5 Video Quality Project: GTE Labs Analysis" analyzed the results of the subjective test. The results are summarized in Tables 1 and 2. Some of the conclusions reached about the methodology of the test based on the results are:

- There was no significant difference between the results produced at the three labs, indicating that there was good control over environmental conditions and recruitment of observers.

- There was no significant difference in the results based on simple observer demographics: age, sex, occupation, visual acuity, or teleconferencing experience.

- The standard error for the mean opinion score of the average HRC-scene combination was only 0.11 on the 1 to 5 scale, which is low enough to serve as a reliable reference for objective measures.

- The mean opinion scores of HRCs correlates well with the video bit rate of the HRCs.

- The 5-point opinion scale seemed to constrain the response for very good or very poor HRCs.

Table 1
Hypothetical Reference Circuits
Ordered By Mean Opinion Score

HRC	MEAN OPINION SCORE	VIDEO BIT RATE (Kbps)	RESOL.	CODING MODE	NOTES
1	4.87	-	-	None	Null codec
3	4.77	45,000	Very High	Proprietary	
2	4.43	-	VHS	None	Baseline
24	3.86	1478	CIF	H.261	Diff.
10	3.79	1536 tot.	High	Proprietary	
25	3.71	1478	CIF	H.261	
22	3.65	710	CIF	H.261	Diff.
8	3.47	768 tot.	Medium	Proprietary	
23	3.30	710	CIF	H.261	Errors
20	3.27	326	CIF	H.261	
9	3.25	768 tot.	High	Proprietary	
7	3.17	384 tot.	Medium	Proprietary	
5	2.96	336 tot.	High	Proprietary	VQ
19	2.76	190	CIF	H.261	Errors, 15 fps
21	2.73	326	CIF	H.261	Errors
18	2.51	118	CIF	H.261	
4	2.38	128 tot.	Medium	Proprietary	VQ
17	2.28	78	CIF	H.261	Diff.
6	2.26	112 tot.	Medium	Proprietary	
14	2.07	326	QCIF	H.261	Diff.
16	1.91	70	CIF	H.261	
13	1.90	118	QCIF	H.261	
12	1.86	70	QCIF	H.261	10 fps, no MC
11	1.83	70	QCIF	H.261	Diff.
15	1.82	62	CIF	H.261	

NOTES:
Unless indicated, all H.261 use Inter coding with motion compensation, with FEC on.
'Diff.' indicates that coder and decoder are from different manufacturers.
'tot.' indicates the total bit rate, the video bit rate is not known.
'VQ' indicates Vector Quantization is the compression method.

Table 2
Test Scenes Ordered By Mean Opinion Score

SCENE	MEAN OPINION SCORE	CONTENT CATAGORY	DESCRIPTION
f	3.67	A	vtc1nw - woman reading news story
k	3.65	A	disguy - male announcer
a	3.65	B	vtc2mp - woman standing next to map
l	3.59	A	disgal - female announcer (Miss Amer.)
u	3.51	D	filter - schematic on pad, pointing
g	3.34	C	5row1 - 5 persons sitting in a row
j	3.25	A	susie - woman talking on phone
b	3.15	B	vtc2zm - woman & map, zoom and pan
p	3.12	C	3twos - 3 pairs of people, scene cuts
w	3.08	B	vowels - teacher at whiteboard
x	3.07	D	inspec - woman at viewgraph projector
d	3.01	C	3inrow - 3 men at table, pan
e	2.80	B	boblec - man lecturing at chalkboard
o	2.76	C	intros - VTC introductions with pans
r	2.72	C	split6 - VTC 3 over 3 format
v	2.69	D	ysmite - map of Yosemite, hand motion
c	2.67	D	washdc - street map, hand motion
y	2.66	E	fredas - Fred Astair tap dancing, B&W
h	2.65	E	flogar - flower garden with pan
t	2.57	D	rodmap - road map with hand motion, pan
m	2.54	B	smity1 - salesman displaying box
n	2.50	B	smity2 - salesman displaying magazine
q	2.31	C	2wbord - 2 people at whiteboard, cuts
s	2.22	D	cirkit - circuit board with zoom
i	2.09	E	ftball - football action

NOTES:
Content categories

- A One person, mainly head and shoulders
- B One person with graphics and/or more detail
- C More than one person
- D Graphics with pointing
- E High object and/or camera motion (broadcast TV)

See Reference B for still images from these test sequences.

3.2 Subjective Test Standards

Work is underway on the six standards listed below which contribute to the subjective assessment of the performance of digital video Teleconferencing systems. Two standards are being developed the ANSI/T1A1.5 (American National Standards Institute), and four are being developed by the ITU (International Telecommunications Union).

ANSI
- T1.801.01 - Subjective Test Scenes
- T1.801.02 - Terms and Definitions

ITU
- P.910 - Subjective Video Quality Assessment Methods for Multimedia Applications
- P.920 - Interactive Test Methods for Audiovisual Communications
- P.930 - Principles of Reference Impairment System for Video
- P.AVQ - Subjective Assessment Methods for Global Audiovisual Quality Evaluation in Multimedia Applications.

3.2.1 Subjective ANSI Standards

ANSI is in the process of finalizing two standards which are related to the subjective assessment of the performance of digital video teleconferencing systems; (1) T1.801.01 - Digital transport of Video Teleconferencing/Video Telephony Signals-Video Test Scenes for Subjective and Objective Performance Assessment, (2) T1.801.02 - Digital transport of Video Teleconferencing/Video Telephony Signals-Performance Terms, Definitions, and Examples. These two standards are briefly described below.

ANSI T1.801.01-TEST SCENES

This standard defines a set of 25 test scenes which were assembled by the ANSI/T1A1.5 committee for use in the development of techniques to assess the performance of digital video teleconferencing and videphone systems. These scenes were used in the subjective tests described in section 3.1, and may be used in future objective tests. The purpose of the standard is to make available to the industry a collection of video test scenes to aid in the development of objective test methodologies that statistically correlate to the subjective performance.

The video scenes have been grouped according to five scene content categories; (1) one person, mainly head and shoulders-four scenes, (2) one person with graphics and/or more detail-seven scenes, (3) more than one person- six scenes, (4) graphics with pointing- five scenes, (5) high object and/or camera motion-three scenes. A single image from each of the test scenes is included in the standard.

All of the video test scenes described in the standard are in the public domain.

Copies of a video tape containing the test scenes are available from ANSI. An Annex describes the timing of the test tape in detail. The typical duration of one test sequence is 13 seconds. A copy of the standard is included in Appendix A.

ANSI T1.801.02-PERFORMANCE TERMS, DEFINITIONS, AND EXAMPLES

This standard specifies a set of terms and definitions which are applicable to the digital transport of Video Teleconferencing/Video Telephony (VTC/VT) signals. The purpose of the standard is to define a common terminology for use in the VTC/VT community, thereby improving communication among current and future members. A total of 15 general terms (e.g. lip sync, scene cut, spatial performance, etc.) and 16 impairment terms (e.g. blurring, edge busyness, jerkiness, etc.) are defined.

A set of illustrative video clips is included with this standard, primarily focusing on the impairment terms but illustrating some of the general terms as well. The tape, which is available from ANSI, is over 10 minutes in length, and contains 26 examples of terms. When illustrating an impairment term, usually both the unimpaired ad impaired versions of the video clip are present on the tape for comparison. A copy of the standard is included in Appendix B.

3.2.2 Subjective ITU Standards

The ITU (Study Group 12) has "determined" that Recommendations P.910, P.920, and P.930 are stable, and it is expected that they will be favorably "decided" at the next Study Group meeting. P.AVQ is a Recommendation which is still is the early drafting stage. Each of the Recommendations is briefly described below.

P.910

Recommendation P.910 (entitled "Subjective Video Quality Assessment Methods for Multimedia Applications") defines non-interactive subjective assessment methods for evaluating the quality of digital video images coded at low and medium bit rates (up to 2Mbit/s) for applications such as videotelephony, videoconferencing and storage and retrieval. The Recommendation covers the following topics.

- Laboratory set up to produce test sequences.
- Laboratory set up to carry out subjective assessment.
- Characteristics of the test sequences.
- Test methods and experimental designs.
- Analysis of data.

This Recommendation does not cover topics that are already included in other Recommendations such as:

- Video reference conditions, defined in the ITU-T draft Recommendation P.930

- Procedures for monitor alignment, described in the CCIR Report 1221
- Interactive test methods, defined in ITU-T draft Recommendation P.920

P.920

Recommendation P.920 (entitled "Interactive Test Methods for Audiovisual Communications") defines interactive evaluation methods for quantifying the impact of coding artifacts and transmission delay on point to point or multipoint audiovisual communications. This methodology is based upon conversation opinion tests, and can be considered to be an extension of methods defined in ITU-T Recommendation P.80, Annex A. Three major topics are covered by the Recommendation; (1) conversational tasks to be performed by test subjects, (2) methods and experimental design, (3) questionnaires.

Interactive audiovisual tests require the definition of a task to be performed by the test subject. The task must be as natural as possible. However, at the same time, it must stimulate the interactive communication, and the outcome must be somewhat quantifiable. Examples of tasks are (1) name-guessing, (2) story comparison, (3) picture-comparison. Several evaluation scales are offered to evaluate the performance of the task.

P.930

Recommendation P.930 (entitled "Principles of a Reference Impairment system for Video") describes the principles of an adjustable video reference system that can be used to generate the reference conditions necessary to characterize the subjective picture quality of video produced by compressed digital video systems. A Reference Impairment System for Video (RISV) can be utilized to simulate the impairment resulting from the compression of video sequences, independent of compression scheme.

An RISV is capable of producing the following categories of distortions, either singly or in combinations, with independent adjustment of each impairment level:

a) Artifacts due to conversions between analog and digital formats (e.g., noise and blurring).
b) Artifacts due to coding and compression (e.g., jerkiness, edge busyness, and block distortion).
c) Artifacts due to transmission channel errors (e.g., errored blocks).

In this recommendation five types of impairments (block distortion, blurring, edge busyness, noise, and jerkiness) are defined and general methods for implementing these impairments are provided. Appendix I describes a specific implementation of these impairments. Other impairments are the subject for future study.

From the viewer=s point of view, the impairments produced by the RISV should

be a good approximation of impairments generated by digital video coding and transmission systems.

Three possible applications for the RISV are: (1) creating reference conditions in subjective tests of digital video systems to ensure that the quality of the scenes presented to viewers covers the entire range of picture quality, (2) defining standard video impairment levels that can be used to compare subjective test results, and (3) quantifying the user-perceived quality of a video system with respect to a known reference.

Although this recommendatioin describes the principles of an RISV, before an implementation can be recommended, validation tests are required.

Appendix I describes VIRIS (a Video Reference Impairment System) developed by Bellcore, which is a specific implementation of an adjustable reference impairment system for video. Although the studies done at Bellcore were with MPEG1, VIRIS can also be used with other compression schemes, such as H.261.

P.AVQ

Draft Recommendation P.AVQ (entitled "Subjective Assessment Methods for Global Audiovisual Quality Evaluation in Multimedia Applications") defines non-interactive subjective assessment methods for evaluating the global quality of digital video with accompanying audio for audiovisual applications such as videotelephony, videoconferencing and storage/retrieval.

The Recommendation describes the methods which are suitable for evaluating the following quality aspects:

1) the global effect of coding impairments, taking into account the interactions between audio and video degradation,
2) problems related to lack of synchronization between the two signals,
3) the impact of transmission errors on such video systems.

These methods can therefore be used for several different purposes including, but not limited to, selection of algorithms, ranking of audiovisual system performance, and evaluation of the quality level during an audiovisual connection.

The Recommendation describes, test methods/experiment design, evaluation procedures, and statistical procedures for reporting results.

3.3 Objective Test Standards

ANSI is actively working on two standards which contribute to the objective assessment of the performance of digital video teleconferencing systems. The first standard (T1.801.03-Digital Transport of One-Way Video Signals; Parameters for Objective Performance Assessment) is near completion in the approval process. The second standard (Visual Channel Delay and Frame rate Measurement) is very early in drafting process. Brief descriptions of the two standards are included below.

T1.801.O3 - PARAMETERS FOR OBJECTIVE PERFORMANCE ASSESSMENT

This standard covers the operational assessment of one-way, 525 line video systems utilizing digital transport facilities. It gives the measurement parameters which may be used to detect changes in the current status of a system when used in comparison with a set of reference measurements on the same system made under initial provisioning circumstances. Additionally, there are diagnostic parameters identified in this standard that may be utilized to characterize aspects of one-way video signals.

This standard specifies methods of measurement of the end-to-end transmission quality (analog input/analog output) for one-way video transmission service channels that employ digital transport. The initial application for this standard is detecting the continued operational readiness of one-way, 525 line video systems utilizing digital transport facilities.

The purpose of this standard is to assure the uniform application of, provide a framework for, and provide definitions of standard video performance parameters for one-way video signals transported digitally by portions of the telecommunications network. This standard is intended to be especially useful as a basis for comparing the present operational readiness of a system with the same system=s past performance. This standard is intended to provide a common understanding by manufactures, carriers, and their customers.

ANSI Committee T1 recognizes that, in the first applications of this standard, accuracy may depend heavily on the expertise and objectivity of the technical staff performing the various measurements. It is expected that, over time, industry experience in applying the standard will reduce the need for a highly-trained staff to correctly apply the standard.

While the subjective tests to date have been probably more extensive than any previously reported in this area, it is clear that any proposed framework and supporting set of key parameters will greatly benefit from initial results reported as a result of real-world application. This framework must be applied extremely carefully; it is believed that its most effective initial use is in the monitoring of changes in performance of installed systems. However, such applications must be critically followed to assess the degree to which results 1) are reproducible, 2) measure changes in systems that result

in visible artifacts and/or distortions, and 3) are insensitive to video sequences different from those used in the T1A1 test program.

It is expected that work will continue to refine and improve the framework and the specification of the key parameters. Therefore, it is expected that this standard will be reissued from time to time.

The standard defines four objective parameters using artificial test signals (e.g. frequency response, active video area, etc.) and fifteen parameters using natural test scenes (e.g. max. added motion energy, % repeated frames, max. added edge energy, etc.). The standard provides an overview of the objective parameters and defines the method of measurement in detail.

VISUAL CHANNEL DELAY AND FRAME RATE (DFR) MEASUREMENT

The ANSI/T1A1.5 standards organization has initiated a project to measure the delay and the transmitted frame rate of a digital video teleconferencing system. Delay and frame rate are key parameters to the assessment of the performance of audiovisual communication system. If the delay is too great, or the frame rate too low, performance suffers. In addition, it is important to measure video delay to determine the preferred level of audio delay to achieve lip synchronization.

The fundamental procedure which is used in this measurement process requires the matching of video frames by the computation of the mean-square-error of a frame pair. In this way input/output frames are compared to measure delay, and a sequence of output frames are correlated with each other to distinguish between transmitted and repeated frames. The process is very computation-intensive, and alternatives are being investigated. One possibility is to insert small artificial reference marks in the each input analog video frame to assist in the spatial and temporal alignment of the output pictures.

3.4 Additional Objective Performance Assessment Studies

The purpose of this section is to discuss additional work which has been, and continues to be, directed toward the development of objective measures of VTC/VP system performance. Work in this area is divided into two parts; (1) development of new artificial test signals, (2) spatial/temporal reference marks. The efforts in these two areas are discussed below.

DEVELOPMENT OF NEW ARTIFICIAL TEST SIGNALS

The performance of analog video systems is accurately defined by quantitative objective standards which are based solely on artificial test signals such as sine waves, impulses, square pulses, color bars, etc. These quantitative objective tests are simple to perform because cost effective test equipment is readily available in the marketplace. There is some hope that, in a similar way, the performance of digital

video systems may also be objectively measured using artificial test signals. To make progress in this direction, work has been undertaken to devise artificial test signals for this purpose.

Objective tests for digital videoconferencing systems are divided in two categories; those which measure distortion in still scenes, and those which measure distortion in moving scenes. It is relatively straightforward to devise artificial signals to measure distortion for stationary scenes. In fact, four objective tests using stationary artificial test signals are defined in ANSI standard T1.801.03 which is designed to test digital videoconferencing systems. Since distortion is frequently evident in most videoconferencing systems when there is a lot of moving detail, it is particularly important to develop objective test for moving scenes. It is this type of test which is very new, having no counterpart in the previous analog test standards. Consequently, it is in this area (measurement of motion distortion) that much of the work in devising artificial test signals has been focused.

Consideration has been given to two different types of artificial test signals to measure motion destortion; the first occupies the entire video frame while the second occupies a window in a natural scene. Delta Information Systems has examined two different types of full-frame test signals designed to measure motion distortion--(1) rotating wheel to measure temporal frequency response, (2) switched dot pattern to measure scene cut response.

The rotating wheel test signal features three different spoke widths, which combine with various rotation speeds to produce 23 different patterns. These patterns are included with the 25 natural test scenes used for the ANSI subjective tests discussed in section 3.1. Preliminary tests have been performed using the rotating wheel, and the results have been inconclusive. It was found to be difficult to develop a reliable test procedure to deal with the appearance of aliased output test patterns.

It was proposed to measure the Scene Cut response (SCR) of a VTC/VP system by the use of switched dot patterns featuring three different sized white dots alternating with a black background. The SCR is defined as the number of frames required for full amplitude response following a pattern switch. The SCR was found to be a poor predictor of subjective performance because video encoders employ widely different coding strategies when a sudden scene change occurs, such as a scene cut. One manufacturer may feel that the user prefers to see the scene continuously build up from the original image to the new image. Another vendor may feel that the user would be distracted by the dynamically changing, distorted transitional images and would prefer to take a long time to transmit the first new frame with very high quality. In both cases, the input scene cut is distorted; it is merely a matter of the type of distortion which different vendors consider least disturbing to the eye. It may be possible to develop an SCR test which is sufficiently robust that it would be able to yield a single universal measure of distortion to an input scene cut regardless whether the codec created a transition or delay distortion.

Consideration has been given to other types of full-frame artificial test scenes. The scene could be composed of multiple objects (e.g. ellipses having differing shapes) moving in a variety of patterns and speeds. The objects could take on a few different characteristics -- uniform/non-uniform brightness, color, etc. The picture background could also be varied in texture according to an easily defined algorithm. In all cases, the objective would be to create rich complex test patterns with relatively simple algorithms so that they can be easily created at the receiver to use as a reference to measure distortion. A key objective is to define dynamic test patterns which will give rise to cleanly identifiable distortions which are typical in VTC/VP systems.

SPATIAL/TEMPORAL REFERENCE MARKS

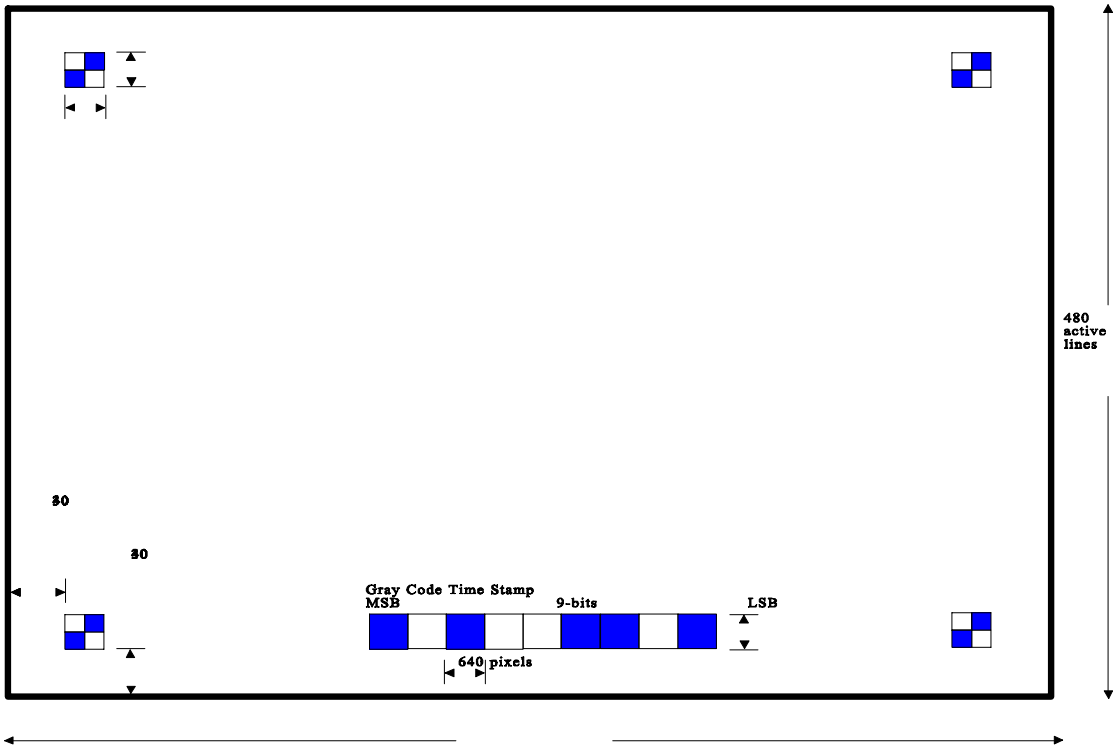
In the execution of objective tests for videoconferencing systems, it is necessary to perform independent measurements in the two fundamental dimensions of the television signal--[1] spatial-within a picture, [2] temporal-from picture to picture. To perform these measurements it is important to be able to align independent TV frames to perform intraframe tests, and to uniquely mark adjacent frames to perform tests in the interframe domain. Delta has initiated a project to insert unique reference marks into any source video signal to accomplish these objectives.

The proposed spatial and temporal reference marks are illustrated in Figure 3.1. The spatial reference marks, also known as fiducial marks, are located in the corners of the image. They occupy 30x30 pixels and are located 40 pixels from the corner of the picture. These marks will make it easy to align two pictures to make measurements such as mean-square-error.

The time code is a nine-bit Gray code where each bit is a 30x30 pixel square. It is centered between the bottom two fiducial marks. The code will continuously increment through 512 unique words and then repeat. In this way, all the frames in a video sequence of 17 seconds length will be uniquely identified. This will greatly simplify the measurement of video delay and frame rate.

Delta has generated the spatial and temporal reference marks and has inserted the pattern in the standard test scene shown in Figure 3.2 (scene M, Smity1, of the scenes used in the ANSI subjective tests). Work is underway to determine whether the proposed code is sufficiently robust and has a sufficiently small impact on the statistics of the original test scene. Preliminary results appear very promising.

LAYOUT OF FIDUCIAL MARKS AND TIME CODES
FOR VTC QUALITY ASSESSMENT



4 CONCLUSIONS AND RECOMMENDATIONS

It has been established that it is very important to improve the procedures to subjectively and objectively assess the performance of digital video teleconferening systems. It has been shown that work to measure video performance by subjective means is relatively mature while the work to achieve this goal by objective procedures is just beginning. It is planned to continue the performance assessment work in the following specific areas.

- continue to support the domestic and international standardization work by the ANSI and ITU organizations in their development of standards to subjectively and objectively assess the performance of multimedia terminals.
- develop, and evaluate, artificial test signals which show promise of objectively measuring the performance of digital video conferencing systems.

5 BIBLIOGRAPHY

1. AThe Effect of Audio-Video Desynchronization on Communication Efficiency in Video Telephony.@ TNO Defense Research, Netherlands.

2. Draft ANSI Standard T1.801.01 AVideo Test Scenes for Subjective and Objective Performance Assessment.@

3. Draft ANSI Standard T1.801-02 AVideo Teleconferencing Performance Terms, Definitions, and Examples.@

4. Draft ANSI Standard T1.801.03 AParameters for Video Teleconferencing Objective Performance Assessment.@

5. Draft ITU-T Standard P.910, ASubjective Video Quality Assessment Methods for Multimedia Applications.@

6. Draft ITU-T Standard P.920, AInteractive Test Methods for Audovisual Communications.@

7. Draft ITU-T Standard P.930, APrinciples of a References Impairment System for Video.@

8. Draft ITU-T Standard P.AVQ, ASubjective Assessment Methods for Global Audio Visual Quality Evaluations in Multimedia Applications.@

9. CCIR 500-5, AMethod for the Subjective Assessment of the Quality of Television Pictures.@

APPENDIX A

APPENDIX B